

Risk Stratification and Prediction of Postoperative Complications Using Temperature Trajectories

R. Padman
Carnegie Mellon University
rpadman@cmu.edu

J. Grant
NorthShore University
HealthSystem
jgrant@northshore.org

U. Ravichandran
NorthShore University
HealthSystem
uravichandran@northshore.org

R. Bashyal
NorthShore University
HealthSystem
RBashyal@northshore.org

N. Shah
NorthShore University
HealthSystem
nshah2@northshore.org

Abstract

Early identification of patients at highest risk of postoperative complications can facilitate appropriate diagnostic work-ups and earlier interventions. We investigate whether postoperative temperature trajectories can stratify patients and predict this risk via a retrospective study of 5,084 adult patients undergoing elective primary total knee arthroplasty (TKA) at a major health system. Demographics, surgery duration, temperature readings, length of stay, comorbidities and complications were extracted from the data warehouse. Group-based trajectory modeling was applied to cluster patients into distinct groups following similar progression of maximum temperature over four-hour time intervals until discharge, and group information was included in predicting risk of critical complications. Three non-overlapping, temperature-based trajectories were identified as high- (8% of patients), medium- (49%), and low-risk (43%) groups. Complication rates were significantly higher in the high-risk group (16.7%), than the medium-risk (5.4%), and low-risk groups (2.70%) ($p < 0.01$). Group information shows promise in improving complication risk prediction for high-risk patients.

1. Introduction

Predictive analytics that leverage clinical data present an encouraging direction for the identification of patients at increased risk of adverse outcomes, especially for determining the type and timing of proactive interventions that may be cost-effective [1, 2]. Postoperative complications and subsequent

readmissions and other adverse events are a significant burden on patients and health systems. Millions of surgeries are performed annually in the US, with surgical hospitalization costs rising to more than 50% of all inpatient costs for the healthcare system [3]. Complication rates range between 6%-47% and average cost per complication ranges from a few hundred dollars for a urinary tract infection to over \$50,000 for ventilator-associated pneumonia, driving increasing healthcare expenditures [4 – 6]. Moreover, postoperative patients without complications who are worked up unnecessarily incur significant morbidity and costs [7-9]. Successful early identification of at-risk patients may lead to timely and accurate workups to reduce adverse events and costs. The current, widely used, postoperative surgical risk stratification tool based only on preoperative covariates lacks capabilities to include some surgery-related features and longitudinal postoperative measurements, such as the development of fever in the days following surgery [10].

When a patient's temperature fluctuates significantly over a short span of time, early identification and mitigation of *individual* risk of adverse events such as infection is particularly challenging. However, *population* risk stratification into distinct groups, such as high, medium and low temperature trajectory groups that best represent the unobserved heterogeneity in patients' postoperative physiology may enable such risk assessment for early intervention. This approach may also provide a better understanding of the dynamic developmental trajectory of fever in the population over the duration of postoperative care.

In this study, we apply group-based trajectory modeling (GBTM) to identify latent clusters of individuals following distinctive trajectories of temperature progression. GBTM is a specialized application of finite mixture modeling applied to longitudinal data to determine a finite number of latent groups which cluster individuals following similar trajectories of the study variable, in this case temperature [11]. While the initial development of GBTM was in the field of criminology, this method has since been applied to study many phenomena, including healthcare [12]. This methodology allows for identification and visualization of the typical trajectory patterns of an outcome of interest and provides a means for risk stratification [2]. We hypothesize that identifying and visualizing postoperative temperature patterns will aid the clinician in the risk assessment of postoperative patients at the point of decision making for appropriate interventions.

Additionally, we leverage the latent class information from GBTM for predicting risk of complications in this patient population. To the best of our knowledge, GBTM has not been applied to study longitudinal temperature modeling in any clinical setting or for predicting risk of postoperative complications. These methods are generalizable to many scenarios where repeated information is collected on individual behaviors [2].

2. Background

Fever is one of the most common symptoms of illness and is typically associated with infection yet remains diagnostically challenging due to the abundant potential etiologies [13]. This is particularly true in the postoperative setting, where a postoperative fever, as defined by a body temperature ≥ 100.4 °F (38°C), may be a normal physiologic response to the trauma of surgery or indicate an underlying pathology such as infection, a thromboembolic event, or hemorrhage [14-16]. The presence of postoperative fever often triggers additional testing which may be costly and unnecessary [17]. While a single abnormal temperature reading may not provide insight into whether or not an infection is present, certain characteristics of postoperative fever may indicate higher risk of a true infection, including magnitude and duration of the fever, and timing of fever onset in relationship to surgery [18]. Robust methods to identify postoperative fever patterns that take into account these fever characteristics may enable better definition of ‘typical’ fever trajectories after surgery

and help clinicians identify patients who may develop a postoperative complication.

Fever after surgery is common, with estimates of postoperative fever ranging from 10% to 40% depending on the type of surgery and definition of fever [18-20]. However, only about 4%-18% of these febrile patients are found to have an infectious etiology [5, 6]. Additionally, infectious complications can occur in patients without postoperative fevers [9]. As a result, the measurement of fever alone is often not sufficient to differentiate postoperative infection and ‘normal’ postoperative physiology. Despite this, diagnostic testing for infections are performed in roughly 40%-60% of febrile postoperative patients [14], resulting in many unnecessary diagnostic evaluations that are costly, potentially invasive and have been shown to be low yield [21]. Improving clinicians’ ability to target only those patients at highest risk of complications, who can benefit from personalized management, has the potential to improve outcomes, value and quality of care.

3. Methods

3.1. Group-Based Trajectory Model (GBTM)

GBTM is an application of finite mixture modeling. Model parameters are estimated using maximum likelihood estimation [11]. The concept of interest is the distribution of the trajectory of a specified outcome, such as postoperative temperature, conditional on a time-related metric, such as time from the completion of surgery. The distribution of outcome trajectories is denoted by $P(Y_i|Z_i)$, where the random vector Y_i represents individual i ’s longitudinal sequence of outcomes and the vector Z_i represents characteristics of i measured at baseline or are otherwise invariant. GBTM assumes that the population distribution of trajectories arises from a finite mixture of unknown order J . The likelihood for each individual Y_i , conditional on the number of groups J , is given by $P(Y_i | Z_i) = \sum_{j=1}^J \pi_j \cdot P(Y_i | Z_i, j; \beta_j)$, where π_j is the membership probability in group j . The conditional distribution of Y_i given membership in j is indexed by the unknown β_j that determines the shape of the group-specific trajectory. The trajectory is modeled using a polynomial function, up to cubic order, of the time-related metric; i.e., $\beta_{0j} + \beta_{1j}t + \beta_{2j}t^2 \dots$. For given j , the method assumes conditional independence for the sequential realizations of the elements of Y_i , y_{it} , over the T periods of measurement. Thus, $P(Y_i | Z_i, j; \beta_j) = \prod_{t=1}^T p(y_{it} | z_{it}, j; \beta_j)$, where $p(\cdot)$ is the distribution of

y_{it} conditional on membership in group j and the z_{it} . A generalization of $P(Y_i|Z_i)$ specifies π_j as a multinomial logit function of baseline characteristics included in Z_i . This allows examination of whether and how such baseline characteristics influence the probability of an individual following the various baseline trajectories [11].

The number of groups that best represents the heterogeneity in the developmental trajectories of the population is obtained by estimating models across a range of trajectory groups and polynomial functions, and comparing their results based on criteria such as the Bayesian Information Criterion (BIC), average probability of group membership and odds of correct classification [11, 22]. The outputs of the model include trajectories that are expressed as polynomial functions of time and probabilities of each individual's posterior probability of being in each trajectory. The trajectory membership for each individual is decided by the individual's highest posterior probability of group membership. For each individual i , the posterior probability of group membership for group j , $PPGM_{ij}$, is calculated using Bayes' Rule as $PPGM_{ij} =$

$$\frac{\pi_j P(Y_i|Z_i, j) \beta_j}{\sum_j \pi_j P(Y_i|Z_i, j) \beta_j}, \text{ based on the complete sequential}$$

realization of the elements of y_{it} over the T periods of measurement. Bayesian updating is used to compute $PPGM_{ij}$ for each time period in between such that the earliest time period when individuals are assigned to their ex-post trajectory group can be determined.

Nonrandom attrition of individuals in different trajectory groups is another modeling issue to address in longitudinal studies since differential attrition rates produce biased parameter estimates of trajectory group sizes and membership probabilities, and influence population-level projections [23]. Dropout due to factors such as death have been modeled using the binary logit function [23]. In the postoperative environment, differential attrition occurs due to patients recovering from surgery at different rates and subsequently being discharged from the hospital, with patients in lower risk trajectories likely recovering faster and receiving earlier discharge.

3.2. Patient Population

This retrospective, observational cohort study includes patients aged ≥ 18 years undergoing unilateral, primary, elective total knee arthroplasty (TKA) at a 4-hospital integrated health system, between January 1, 2007, and December 31, 2013. Patients were excluded if past medical history was

missing, if the surgical start and end times were missing or conflicting, if patients were undergoing revision TKA or if pre-operative infection was suspected and microbiologic cultures were collected prior to the end of TKA. Patients were only included once, thus any data from patients with subsequent TKA during the study period were excluded. This study was approved by the Institutional Review Board of the health system.

3.3. Clinical Protocols

All TKA were performed using regional anesthesia unless contraindicated or unsuccessful. Anesthesia included a single-shot spinal block, sedation, and use of additional peripheral nerve block as needed. Antibiotic prophylaxis was routinely given to all patients one hour prior to incision and subsequently continued for 24 hours. Patients were routinely admitted to the hospital and monitored after TKA, with temperature collected every 4 hours. While fever was defined as any temperature $\geq 100.4^\circ\text{F}$, we also examined temperature $\geq 101^\circ\text{F}$.

3.4. Data Collection

Data were extracted from the health system's Enterprise Database Warehouse (EDW) for this study. We identified patients using Current Procedural Terminology (CPT) and institution-specific procedure codes for TKA. Data included age, sex, race/ethnicity, body-mass index (BMI), medical co-morbidities, temperature measurements and site of temperature measurement, medications received and time administered, start and end time of surgery, discharge diagnoses (ICD-9 codes), length of stay (LOS), hospital readmissions and death. We obtained patient data from the index TKA hospitalization and any subsequent encounters within 30 days of surgery. We assessed each demographic variable individually for consistency and potential inaccuracies, such as removing outliers (age < 18 , conflicting surgery records, etc.). When feasible, missing data was handled via reasonable imputation such as recording missing BMI as the average for the gender and age group of the patient.

The EDW was queried for complications based on ICD-9 codes for the following conditions: pneumonia, urinary tract infection (UTI), cholecystitis, bacteremia, skin and soft tissue infection (SSTI), surgical site infection (SSI), deep vein thrombosis (DVT), pulmonary embolism (PE), cerebrovascular accident, medication reaction, acute blood loss or hematoma and transfusion reaction. Bacteremia was

corroborated by positive blood culture. Acute blood loss and hematoma was determined if a patient had a hemoglobin difference of 5 g/dL or ≥ 3 units of blood transfused during index hospitalization. *Clostridium difficile* infection was determined by positive *Clostridium difficile* enzyme immunoassay or polymerase chain reaction tests. Death date was determined from the Social Security Death Index. A physician reviewed each chart identified as having a complication to verify the diagnosis and the date that symptoms started by assessing subjective complaints, vital signs, consistent imaging findings, microbiologic data, corroborating laboratory data (i.e., white blood cell count), documentation in the chart of the complication and actions taken by the treating team (i.e. starting antibiotics).

3.5. Modeling the Fever Trajectories

Temperature measurements were categorized into four-hour time periods after surgery for up to five days and the maximum oral temperature, T_{max} , in each period was used to model the trajectories using a censored normal distribution with limits of 89°F - 105°F. All non-oral temperatures were excluded from analysis. Furthermore, any missing temperatures can be handled by GBTM intrinsically, and further justifies using this approach in this study.

Group-based trajectory analysis was performed in four steps. The first step identified the optimal baseline trajectory using only temperature data. In the second step, covariates considered relevant for TKA (i.e. gender, age, BMI, procedure time) were added to estimate the effect of individual risk factors measured at baseline on the probability of trajectory group membership. The third step adjusted the parameter estimates for nonrandom attrition of patients due to differential discharge-on-recovery by adding a binary logit model with a length of stay (LOS) criterion, where $LOS > 3$ days was used as an indicator of patients who may be at higher risk of complications due to elevated temperatures, and hence stay longer in the hospital. The time period for trajectory analysis was selected to be 2 days longer than the average LOS of the population because more than 85% of the patient population is discharged by 5 days after TKA surgery. Finally, in the fourth step, $PPGM_{ij}$ was computed for each patient's posterior probability for membership in each of the trajectory groups, j , in each time interval using their longitudinal sequence of T_{max} measurements. This analysis determined the earliest 4-hour period that indicated a patient's final trajectory group.

We used Stata (Version 14.2, StataCorp. College Station, TX) with the command *traj* for estimating

GBTM [24], while descriptive statistics and analysis of model results were performed using R Version 3.3.2 (R Foundation for Statistical Computing, 2014, Vienna, Austria). Patient characteristics and postoperative complications were compared across group trajectories using Student's t-test, Pearson's Chi-Square test and Fisher's Exact test. A p-value of <0.05 was considered statistically significant.

3.6. Complications Prediction

While risk stratification models assign patients to their most likely trajectory groups or latent classes, predicting an individual patient's risk of complications after surgery is the eventual outcome of interest to decision makers. Hence, the latent class information from GBTM is incorporated into several widely used statistical and machine learning models from the literature, such as logistic regression, random forest and decision trees, to classify patients based on their likelihood of developing complications after surgery. The models include preoperative features such as demographic and comorbidity data, surgery duration, and trajectory group information before the first complication is recorded. With only 5.11% of patients experiencing any complication in this low-risk TKA surgery, we performed oversampling to handle the highly imbalanced data, and report F1 score on the test dataset that balances precision and recall metrics, using an 80-20 random split.

4. Results

4.1. Clinical and Demographic Features

6,392 adult patients underwent elective primary TKA. We excluded 873 patients with more than one surgery (typically, contralateral TKA) during the time period, 387 patients with exclusively non-oral temperature readings, 38 emergent surgery patients and 10 patients with missing surgical dates and times, resulting in 5,084 unique patients. Patient characteristics are summarized in Table 1. Overall, 3,324 (65%) patients were female, the mean age was 68.6 (SD=9.9) years, mean BMI was 30.7 (SD=6.4) kg/m² and mean procedure time was 121.41 minutes (SD=47 minutes). With a mean LOS of 3.2 (SD=1.1) days, majority of patients (4269/83.9%) were discharged on or before the 3rd day.

4.2. Postoperative Temperatures

A total of 104,568 temperature measurements were obtained from the EDW for 5,084 patients for up

to five days of inpatient stay (almost 21 measurements/patient), with oral temperature measurements constituting the majority at 87.2% (91,227 readings), followed by tympanic (5.4%), temporal (2.3%), axillary (2.1%), core (0.2%), and rectal (0.04%). 2.8% of the measurements had missing temperature site values. 3,503 (68.9%) patients had a *preoperative* temperature recorded, and of these patients, 3 (0.06%) had a temperature $\geq 100.4^{\circ}\text{F}$. *Postoperatively*, we found 717 (14.1%) patients had a fever ($\geq 100.4^{\circ}\text{F}$) at any point after surgery and 313 (6.51%) had a temperature $\geq 101^{\circ}\text{F}$ (Table 1).

Table 1. Patient characteristics by trajectory group

Characteristic	Overall	Group 1	Group 2	Group 3	P-value
N (%) or mean (SD)	N=5084	N=2189	N=2493	N=402	
Mean age (SD)	68.58 (9.96)	70.86 (9.79)	66.86 (9.69)	66.80 (10.06)	<0.01
Female (%)	3324 (65.38%)	1463 (66.83%)	1619 (64.94%)	242 (60.20%)	0.02
Mean BMI, kg/m ² (SD)	30.70 (6.42)	29.63 (6.01)	31.31 (6.42)	32.81 (6.82)	<0.01
Diabetes Mellitus (%)	713 (14.02%)	238 (10.87%)	391 (15.68%)	84 (20.90%)	<0.01
Rheumatoid Arthritis (%)	275 (5.41%)	140 (6.40%)	112 (4.49%)	23 (5.72%)	0.02
COPD (%)	599 (11.78%)	270 (12.33%)	291 (11.67%)	38 (9.45%)	0.25
CHF (%)	202 (3.97%)	108 (4.93%)	85 (3.41%)	9 (2.24%)	<0.01
Hypertension (%)	2729 (53.68%)	1205 (55.05%)	1292 (51.83%)	232 (57.71%)	<0.01
Mean TPT, minutes (SD)	121.41 (47.23)	123.86 (52.04)	118.37 (40.03)	126.85 (58.42)	<0.01
Mean LOS (SD)	3.21 (1.06)	3.18 (0.98)	3.18 (0.99)	3.59 (1.64)	<0.01
LOS > 3 Days (%)	815 (16.03%)	297 (13.57%)	388 (15.56%)	130 (32.34%)	<0.01
Antipyretic Doses (SD)	13.5 (5.9)	14 (5.8)	13.1 (5.8)	12.5 (6.4)	<0.01
Postoperative fever $\geq 100.4^{\circ}\text{F}$	717 (14.1%)	22 (1.01%)	379 (15.2%)	316 (78.61%)	<0.01
Postoperative fever $\geq 101^{\circ}\text{F}$	313 (6.51%)	3 (0.14%)	129 (5.17%)	199 (49.5%)	<0.01

Abbreviations: BMI, body-mass-index; COPD, Chronic obstructive pulmonary disease; CHF, congestive heart failure; TPT, total procedure time; LOS, length of stay.

4.3. Fever Trajectories

With baseline demographic and clinical covariates and differential attrition based on LOS, GBTM evaluated varying number of trajectory groups using BIC and optimally identified three distinct postoperative temperature trajectory groups following quadratic, cubic and cubic trajectories, respectively (Figure 1). Predicted and actual average T_{max} of each group, in 4-hour increments with non-overlapping 95% confidence intervals around the estimates, are displayed in Figure 1. The low T_{max} trajectory (group 1) includes 2,189 (43%) patients of which 1% experienced a fever, the medium T_{max} trajectory (group 2) has 2,493 (49%) patients of which 15.2% experienced a fever and the high T_{max} trajectory (group 3) includes 402 (8%) patients of which 78.6% experienced a fever. Group 3 patients show the highest average T_{max} throughout the five days and almost 50% reported a temperature $\geq 101^{\circ}\text{F}$.

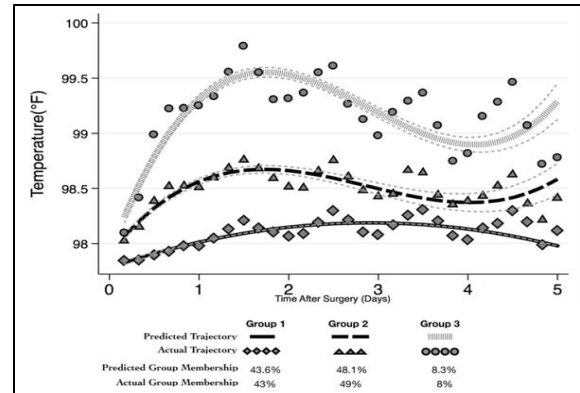


Figure 1. Temperature trajectories

As seen in Table 1, group profiling shows statistically significant differences across all groups. Group 1 patients are older, group 2 has the lowest total procedure time and group 3 patients are more likely to be male, have a higher BMI, have diabetes and hypertension, and have longer procedure times and LOS (Table 1). LOS was incorporated into the model to control for the systematic discharge of afebrile patients; patients with fevers were less likely to be discharged home within three days of surgery. Only 13.6% of patients in group 1 were still hospitalized after 3rd day as compared to 15.6% in group 2 and 32.1% in group 3 ($p < 0.01$).

4.4. Early Identification of High Fever Patients

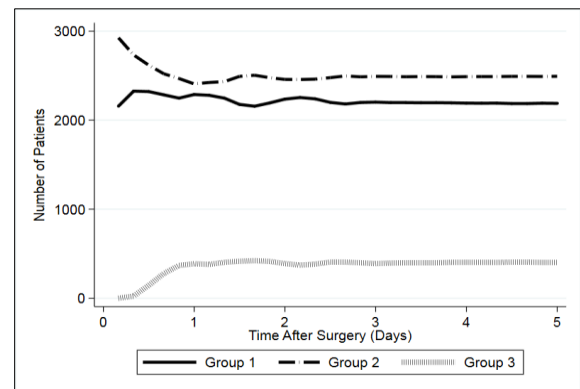


Figure 2. Trajectory convergence graph

Figure 2 depicts the convergence of the trajectory model based on the posterior probability of group membership, $PPGM_{ij}$, to assign patients to one of three groups as new temperature data becomes available. We observe that all three trajectories stabilize within 24-36 hours. Figure 3 demonstrates that almost 70% of patients in group 3 and over 80% of patients in groups 1 and 2 are correctly classified into their final trajectory group within 36 hours.

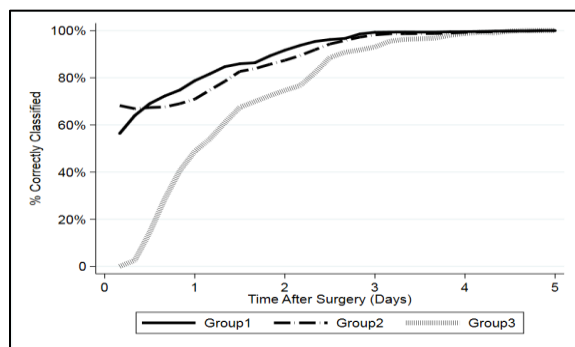


Figure 3. Classification accuracy over time based on final trajectory group

4.5. Predictors of Trajectory Group Membership

Table 2 quantifies whether and how patients' clinical and surgical characteristics, stratified by trajectory group, influence the probability of an individual following the various baseline trajectories. The statistically significant results indicate that, as compared to patients in the lowest temperature trajectory (group 1), age is less likely to be a risk factor for groups 2 and 3 ($p < 0.0001$), while BMI ($p < 0.0001$) and presence of diabetes ($p < 0.001$) are more likely to be so and total procedure time was less likely to be a risk factor for group 2 ($p < 0.001$) and not significant for group 3. Male gender and other co-morbid conditions had no statistically significant impact on membership in a specific trajectory.

Table 2. Covariates of interest

	Group 1	Group 2		Group 3	
		Parameter	P-value	Parameter	P-value
Male		0.025	0.76	0.212	0.086
Age At Discharge (years)		-0.033	<0.0001	-0.027	<0.0001
BMI		0.025	<0.0001	0.056	<0.0001
Total Procedure Time (minutes)		-0.003	<0.001	0.0002	0.881
COPD	Ref.	-0.013	0.92	-0.295	0.142
Diabetes		0.378	<0.001	0.601	<0.001
Rheumatoid Disease		-0.353	0.037	-0.052	0.838
Congestive Heart Failure		-0.255	0.18	-0.722	0.053
Hypertension		-0.074	0.36	0.007	0.954

Abbreviations: BMI, body-mass-index; COPD, chronic obstructive pulmonary disease

We examined the use of antipyretic medications during postoperative care given the direct impact these medications may have on lowering temperature. Antipyretics are given in response to a fever, but also used following surgery for analgesia, often in combination with an opiate. Overall, patients received

a mean of 13.5 (SD=5.9) antipyretic doses across their hospital stay, with a per day average of 4.3 doses (SD=1.8). There were significant differences across groups in the number of doses per patient per day (group 1: 4.5 (SD = 1.8), group 2: 4.2 (SD = 1.7), group 3: 3.6 (SD = 1.5), $p < 0.01$).

4.6. Trajectory Groups and Postoperative Complications

Postoperative complications across the three trajectories are shown in Table 3. We observe higher percentages of all complications in group 3 as compared to groups 1 or 2. Risk stratification was statistically significant for DVT/PE, pneumonia, UTI, SSTI, SSI, bacteremia and bleeding. The most common complication was DVT/PE, occurring in 134 (2.6%) of all patients. There were very low rates (<0.5%) of SSI, *Clostridium difficile* infection, bacteremia and cholecystitis, limiting comparative analysis. *Any complication*, which is a composite of all the infectious and non-infectious complications evaluated, occurred in 16.7% (67/402) of patients in group 3 as compared to 5.4% (134/2493) in group 2 and 2.7% (59/2189) in group 1 ($p < 0.01$).

Table 3. Complication rates by trajectory group

Characteristic	Overall N (%)	Group 1 N=5084	Group 2 N=2493	Group 3 N=402	P-value
Any Complication	260 (5.11%)	59 (2.70%)	134 (5.38%)	67 (16.67%)	<0.01
Bacteremia	1 (0.02%)	00 (0.00%)	00 (0.00%)	01 (0.25%)	<0.01
Bleeding	35 (0.69%)	07 (0.32%)	20 (0.80%)	08 (1.99%)	<0.01
<i>Clostridium difficile</i> infection	2 (0.04%)	01 (0.05%)	00 (0.00%)	01 (0.25%)	0.06
Cholecystitis	2 (0.04%)	00 (0.00%)	01 (0.04%)	01 (0.25%)	0.07
CVA	5 (0.10%)	01 (0.05%)	04 (0.16%)	00 (0.00%)	0.37
Death within 30 Days of Surgery	3 (0.06%)	01 (0.05%)	02 (0.08%)	00 (0.00%)	0.78
Drug Reaction	3 (0.06%)	03 (0.14%)	00 (0.00%)	00 (0.00%)	0.14
DVT/PE	134 (2.64%)	27 (1.23%)	71 (2.85%)	36 (8.96%)	<0.01
Pneumonia	42 (0.83%)	11 (0.50%)	17 (0.68%)	14 (3.48%)	<0.01
SSTI	7 (0.14%)	01 (0.05%)	02 (0.08%)	04 (1.00%)	<0.01
Surgical Site Infection	1 (0.02%)	00 (0.00%)	00 (0.00%)	01 (0.25%)	<0.01
Transfusion Reaction	3 (0.06%)	00 (0.00%)	02 (0.08%)	01 (0.25%)	0.14
UTI	55 (1.08%)	15 (0.69%)	30 (1.20%)	10 (2.49%)	<0.01
Readmission within 30 Days	180 (3.54%)	82 (3.75%)	76 (3.05%)	22 (5.47%)	0.04

Abbreviations: CVA, cerebrovascular accident; DVT, deep vein thrombosis; PE, pulmonary embolism; SSTI, skin and soft tissue infection; UTI, urinary tract infection

4.7. Complications Prediction

Table 4 presents the summary results of the prediction models showing the best model among those evaluated which included Logistic Regression, Decision Trees, K-Nearest Neighbors and Random Forest. Adding

trajectory group information of individual patients improved the F1 score marginally when the entire population was included in the prediction analysis. However, when the trajectory group information of each patient was used to develop a group specific prediction model, we observe improved F1 score in the high-risk group, demonstrating the potential of risk stratification to improve prediction of complications in this critical group and inform better clinical decisions.

Table 4. Prediction model performance

Scenario	Best Model	Performance Metric
No Group Information	Decision Tree	F1=0.40
Group Information	Decision Tree	F1=0.42
Comparison between the three groups		
High risk group	Decision Tree	F1=0.55
Medium risk group	Decision Tree	F1=0.30
Low Risk group	Random Forest	F1=0.2

5. Discussion

We applied GBTM to a large dataset as a novel approach for temperature interpretation, finding three distinct temperature trajectories. Patients with diabetes, higher BMI, younger age, and longer procedure times were more likely to belong to the higher temperature trajectory groups, which is consistent with previous literature [25-26]. We identified a statistically significant increase in the frequency of surgical complication from group 1 to group 3, respectively. However, as evidenced in the literature and practice, TKA is a low complication surgery, hence post-operative risk stratification may need to be combined with pre-operative and surgery related characteristics to potentially be more successful in predicting rare occurrence of post-surgical complications in a large proportion of the patient population. This is a challenge to be addressed in the future as more granular data is collected in pre-operative and surgical settings in addition to the post-operative environment analyzed in this study.

This risk stratification carries implications for both the high and low risk patient populations. Studies have found that patients who experience fever on or after 3rd postoperative day are more likely to have complications [27]. With potentially earlier identification of high-risk patients, clinicians may be able to perform focused, preemptive clinical

monitoring or better targeted evaluations and interventions. Correct convergence of almost three quarters of the patients to their final temperature trajectory in 36 hours highlights the potential benefit of using GBTM for earlier identification of patients at higher risk.

While this approach holds promise, more work will be required to see if this is more accurate and/or faster at determining high risk patients than current methods of clinical assessment, including an elevated temperature, white blood cell count or concerning symptoms. Conversely, another benefit of temperature trajectory modeling and risk stratification is the ability to identify patients early at very low risk of having a postoperative complication. While fever on or after postoperative day 3 is more likely to reflect a complication, fever is most common on day 1 and declines from there. If the temperature dynamics are such that a patient with an early postoperative fever is predicted to belong to a low trajectory group, it may be possible to avoid ordering unnecessary tests or prolonged LOS based on the reassuring temperature trajectory.

In our study, 2189 (43%) patients in the low temperature trajectory had a <3% hospital complication rate. This group may be targeted for early discharge or other cost- and resource-savings strategies with potential downstream improvements in patient outcomes and satisfaction. Therefore, an important next step is a detailed comparison of the status quo process of a cross sectional measurement of fever or no fever vs. the GBTM approach. A single cross sectional measurement of fever or no fever, or alternatively a T_{max} in the last 24 hours, may be an overly simplistic approach at identifying risk. New computational tools integrated into a clinical workflow could allow us to extract much more value out of a traditional measure. Thus, future work will need to demonstrate the benefit of GBTM or other methods to traditional risk analysis based on temperature as a binary variable.

6. Limitations

There are some limitations to the current study. This is a retrospective study that obtained clinical data from an EDW. The EDW may contain inaccurate documentation of vital signs, surgical parameters and clinical co-morbidities, however the large number of patients analyzed in this study may offset small data errors [28]. A physician reviewed all charts with a complication as identified through the EDW, but due to the large volume, negative charts were not reviewed. While this may result in an underestimation

of specific complications, we do not suspect this would alter the underlying risk stratification. An internal framework was developed to determine the presence or absence of each particular complication, but this was not standardized to national postoperative outcomes definitions such as the National Surgical Quality Improvement Project (NSQIP); this is planned for future work. As a retrospective study over a seven-year period, changes in clinical protocols or patient population over time may have an unknown impact on the results. Additionally, while we included age, gender, BMI and various clinical co-morbidities in the model, it is possible that other unobserved clinical conditions affected temperature.

Our evaluation of the impact of medications on temperature trajectories is limited. We noted differences in the use of antipyretic medications between the groups, with group 1 having the highest use of antipyretics. However, with 14 doses, on average, of antipyretic in group 1 vs. 13.1 doses, on average in group 2, with the same standard deviation and over exactly the same average LOS, there may be a question whether patients with higher antipyretic use were ‘falsely’ assigned to lower temperature trajectories. While this needs more exploration, we are reassured that we still see significant risk stratification, and the statistical significance may be driven by cohort sizes. Future work will involve analysis of the time-varying impact of antipyretics which may improve risk stratification. Additional medications such as antibiotics may also impact temperature if an infection is present and will be evaluated in future work. In this cohort, nearly all patients (per protocol) received antibiotics during the first 24 hours of their hospital stay.

Another important limitation is the lack of standardization among temperature measurement sites. It is known that peripheral temperature sites including oral temperature result in cooler temperature readings as compared to core temperature sites. In this study, we removed patients (n=387) with exclusively non-oral temperatures as well as the few non-oral temperature measurements of the remaining 5,084 patients (less than 13%). While the effect on current trajectory group assignment is likely to be small, as the number of non-oral temperature readings increase, in other surgery types, for example, a full sensitivity analysis will need to be performed. Given the large number of variables that can affect temperature from any given measurement site, a correction factor that normalizes different sites has been noticeably absent from the literature.

In contrast to these limitations, significant advantages of the study are the large number of patients included and that surgeries were performed at

four different hospitals within the health system, both teaching and non-teaching institutions, supporting the potential generalizability of these findings. The study also illustrates the potential clinical value of the trajectory of a single metric - temperature measurement - which is routinely collected in every post-surgical setting, to risk stratify patients.

7. Conclusions

We illustrate the potential of GBTM for postoperative temperature modeling and identification of latent risk groups in a large cohort of TKA patients, one of the most common inpatient surgeries in the United States. Using this approach to analyze temperature, we delineated three distinct trajectories of postoperative temperature in this population with early convergence into the final trajectory group that facilitates clinical decision making, characterized the clinical and surgical features of patients in the different fever trajectories, identified increased surgical complication rates from group 1 to group 3, respectively, and predicted the likelihood of complications in this low risk TKA surgery.

This work lends itself to further analysis, including comparison of temperature trajectories of different surgeries, predicting the presence of specific complications, comparative assessment of a patient’s risk of specific complications and evaluating its impact in guiding cost-effective postoperative diagnostic evaluations. Further, this approach could be applied to other vital signs modeled as time-varying covariates or multi-trajectories, predicting more specific outcomes for individual patients, or to improve the sensitivity/specificity of identifying the highest or lowest risk patients for different outcomes [29-30].

This is a first step in the overall research objective to develop the capability for real-time identification of patients who are at highest risk of *any* and *specific* postoperative complications, thus facilitating appropriate and more targeted work-up of postoperative patients and earlier interventions when needed, with the potential to improve quality of care.

8. Acknowledgements

The authors are grateful to the Daniel F. and Ada L. Rice Foundation for their generous contributions to support this work. We also thank Drs. R. Parikh, M. McNulty, A. Robicsek and DS Nagin for inputs on the problem and modeling and graduate students S. Kodi, M. Sharma, M. Turner, V. Vegi, K. Kinnard, P. Rao,

P. Raja and Y. Mathiyalagan who assisted with the data analysis.

9. References

- [1] Wang W, Zhao H, Zhuang H, Shah N, Padman R. DyCRS: Dynamic Interpretable Postoperative Complication Risk Scoring. WWW'20: *Proceedings of the Web Conference 2020*, Taiwan, <https://doi.org/10.1145/3366423.3380253>.
- [2] Ben-Assuli O, Padman R. Trajectories of Repeated Readmissions of Chronic Disease Patients: Risk Stratification, Profiling, and Predictions. Special Issue of *MIS Quarterly on The Role of Information Systems and Analytics in Chronic Disease Prevention and Management*, 2020, 201-226; DOI: 10.25300/MISQ/2020/15101.
- [3] Health Care Cost Institute. *Health Care Cost and Utilization Report*: 2018. Accessed: 13 June, 2021 Available at: <https://healthcostinstitute.org/annual-reports/2020-02-13-18-20-19>.
- [4] Kremers MH, Visscher SL, Moriarty JP, Reinalda MS, Kremers WK, Naessens JM, Lewallen DG. Determinants of direct medical costs in primary and revision total knee arthroplasty. *Clin Orthop Relat Res*. 2013;471(1):206-14. doi: 10.1007/s11999-012-2508-z. PMID: 22864619; PMCID: PMC3528929.
- [5] Walid MS, Sahiner G, Robinson C, Robinson JS, Ajjan M, Robinson JS. Postoperative fever discharge guidelines increase hospital charges associated with spine surgery. *Neurosurgery*. 2011;68(4):945-949.
- [6] Blumstein GW, Andras LM, Seehausen DA, Harris L, Ross PA, Skaggs DL. Fever Is Common Postoperatively Following Posterior Spinal Fusion: Infection Is an Uncommon Cause. *J Pediatr*. 2015 Mar 1;166(3):751-5.
- [7] Yoo J-H, Restrepo C, Chen AF, Parvizi J. Routine Workup of Postoperative Pyrexia following Total Joint Arthroplasty is Only Necessary in Select Circumstances. *Journal of Arthroplasty*, 2016. Available: <http://www.sciencedirect.com/science/article/pii/S0883540316306684>.
- [8] Schey D, Salom EM, Papadia A, Penalver M. Extensive fever workup produces low yield in determining infectious etiology. *Am J Obstet Gynecol*. 2005 May;192(5):1729-34.
- [9] Vermeulen H, Storm-Versloot MN, Goossens A, Speelman P, Legemate DA. Diagnostic Accuracy of Routine Postoperative Body Temperature Measurements. *Clin Infect Dis*. 2005 May 15;40(10):1404-10.
- [10] Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmieciak TE, Ko CY, and Cohen ME. Development and evaluation of the universal acs nsqip surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J of the American College of Surgeons*, 217(5):833-842, 2013.
- [11] Nagin DS. *Group-Based Modeling of Development*. Cambridge, MA: Harvard University Press; 2005.
- [12] Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*. 2010;6:109-38.
- [13] Mackowiak PA. Concepts of fever. *Arch Intern Med*. 1998 Sep 28;158(17):1870-81.
- [14] Frank SM, Kluger MJ, Kunkel SL. Elevated Thermostatic Setpoint in Postoperative Patients. *J Am Soc Anesthesiol*. 2000 Dec 1;93(6):1426-31.
- [15] Ishii Y, et al. Characteristics and significance of fever during 4 weeks after primary total knee arthroplasty. *Arch Orthop Trauma Surg*. 2014 May;134(5):707-12.
- [16] Athanassios C, Samad A, Avery A, Cohen J, Chalmick D. Evaluation of fever in the immediate postoperative period in patients who underwent total joint arthroplasty. *J Arthroplasty*. 2011;26(8):1404-8.
- [17] Ward DT, Hansen EN, Takemoto SK, Bozic KJ. Cost and Effectiveness of Postoperative Fever Diagnostic Evaluation in Total Joint Arthroplasty Patients. *J Arthroplasty*. 2010;25(6, Supplement):43-8.
- [18] de la Torre SH, Mandel L, Goff BA. Evaluation of postoperative fever: Usefulness and cost-effectiveness of routine workup. *Am J Obstet Gynecol*. 2003 Jun;188(6):1642-7.
- [19] C Galicier, et al. A prospective study of postoperative fever in a general surgery department. *Infect Control* 1985.
- [20] C. A. Perlino, Postoperative fever, *Med Clin North Am*, 2001.
- [21] Lesperance R, et al. Early postoperative fever and the "Routine" Fever Work-Up: results of a prospective study. *Journal of Surgical Research* 2011.
- [22] Klijn SL, Weijenberg MP, Lemmens P, van den Brandt PA, Lima Passos V. Introducing the fit-criteria assessment plot - A visualisation tool to assist class enumeration in group-based trajectory modelling. *Stat Methods Med Res*. 2017 Oct;26(5):2424-36.
- [23] Haviland A, Nagin DS, Rosenbaum PR, Tremblay RE. Combining group-based trajectory modeling and propensity score matching for causal inferences in nonexperimental longitudinal data. *Dev Psychol*. 2008 Mar;44(2):422-36.
- [24] Jones BL and Nagin DS. 2013. A Note on a Stata Plugin for Estimating Group-based Trajectory Models. *Sociological Methods Research*, vol. 42(4): 608-613.
- [25] Belmont PJ, Goodman GP, Waterman BR, Bader JO, Schoenfeld AJ. Thirty-Day Postoperative Complications and Mortality Following Total Knee Arthroplasty. *J Bone Jt Surg Am*. 2014;96(1):20-6.
- [26] Norman DC. Fever in the elderly. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2000 Jul;31(1):148-51.
- [27] Lu X, Jin J, Lin J, Qian W, Weng X. Course of fever and potential infection after total joint replacement. *Knee Surg Sports Traumatol Arthrosc Off J ESSKA*. 2015 Jun;23(6):1870-6.
- [28] Fottrell E, Byass P, Berhane Y. Demonstrating the robustness of population surveillance data: implications of error rates on demographic and mortality estimates. *BMC Med Res Methodol*. 2008;8:13.
- [29] Burckhardt P, Nagin DS, Padman R. Multi-Trajectory Models of Chronic Kidney Disease Progression. *AMIA Annu Symp Proc*. 2017:1737-46.
- [30] Burckhardt P, Nagin DS, Vijayasarathy VPR, Padman R. Multi-Trajectory Modeling to Predict Acute Kidney Injury in Chronic Kidney Disease Patients. *AMIA Annu Symp Proc*. 2018:1196-1205.